



**BIOINFORMATICS 2015 FALL SEMINAR SERIES**

Hosted by: Department of Computer and Information Sciences,  
Department of Electrical and Computer Engineering &  
Center for Bioinformatics and Computational Biology  
<http://bioinformatics.udel.edu/Seminars/Current>

**MONDAY, February 15, 2016**  
**3:30pm**  
**DBI Room 102**

## **Vascular Trees and the Machine Learning Strategy on Manifold**

***Hongyu Miao***

***Associate Professor***

***Department of Biostatistics, School of Public Health  
University of Texas HSC***

**ABSTRACT:** In the era of Data Science, availability of complex data in multiscale biology is leading to increasing attempts to develop novel and efficient statistical and computational methodologies. Here we primarily focus on the tree-structured data to illustrate how such data can be quantitatively parameterized and analyzed. More specifically, tree-structured data contain both topological and geometrical information, and are necessarily considered in manifold instead of Euclidean space. Thus, we propose a novel data parameterization method, called Topology-Attribute matrix (T-A matrix), so the statistical learning task can be conducted on matrix manifold. We incorporate the structure constraints embedded in data into the negative matrix factorization method to determine the meta-trees from the forest matrix, and the signature vector of each single tree can then be extracted by meta-tree decomposition. The meta-tree space turns out to be a cone space, in which we explore the distance metrics and implement the unsupervised learning algorithm based on the transition path through the consensus tree. Finally, the T-A matrix based clustering (TAMBAC) framework is evaluated and compared using both simulated data and real retinal images to illustrate its efficiency and accuracy.